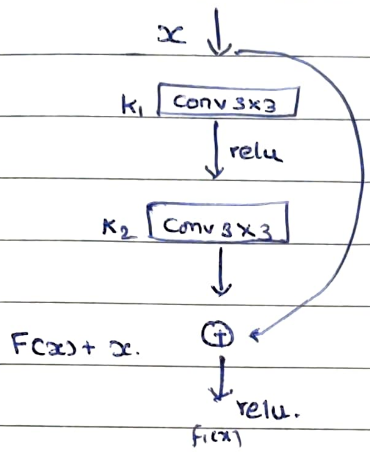
Assignment-3

1) Rescidual-Block :

Let us go with basic residual Block without batch-norm to understand the gravity of the resnet's.



Forward prop :

$h(x) = r(x * K_1)$ (Kernel K_1) $\left[r(x) = \begin{cases} x & x>0 \\ 0 & x<0 \end{cases} \text{(relu function)} \right]$

$F(x) = K_2 * h(x)$

Backprop:

The backprop which we are intrested about is $\frac{\partial f_i(x)}{\partial x}$ as x contains contributions of privious layers.

$$\frac{\partial f_i(x)}{\partial x} = \frac{\partial F(x)}{\partial x} + 1$$

$$= 1 + \frac{\partial F(x)}{\partial x}.$$ (∴ assuming they are "+" ve)

By this process we are ensuring that the gradients are not dying off as they have their contribution of it

2) To find the support the one way is to back track things so one thing let us assume the final pixel is 1x1 image of 4th conv layer.

For nxn image $\xrightarrow{conv}$ $(n-2)\times(n-2)$ image

∴ By Back tracking

$1\times1 \leftarrow 3\times3 \leftarrow 5\times5 \leftarrow 7\times7 \leftarrow 9\times9$

∴ The support is from 81 pixels to this one.

classmate
Date
Page

3) Adding more units increases (may) number of regions that affect the output this indeed decreases the bias as more are contriubuting but increases the variance which can lead to be overfit with the data. A general rule of thumb to remember is if the network becomes complex the bias would decrease but variance would increase

4) Given,

$$\tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}$$

We know that that the relationship between tanh and sigmoid.

$$\tanh(a) = 2\sigma(2a) - 1$$

$$\Rightarrow \sigma(2a) = \frac{\tanh(a) + 1}{2} \Rightarrow \boxed{\sigma(a) = \frac{\tanh(a/2)}{2} + \frac{1}{2}}$$

So By this we can make the weights of input to sigmoid function of ½ and make weight of that neuron ½ and add a ½ to the bias function So this is a linear transformation so that it can be achieved.

Date
Page

5) Given,

$$E(w) \approx E(w^*) + \frac{1}{2}(w-w^*)^T \cdot H(w-w^*)$$

Now we are intrested in contours of constant error and let it be "c"

$$\therefore \quad C = \frac{1}{2}(w-w^*)^T \cdot H(w-w^T)$$

Given that $w - w^* = \Sigma \alpha_i u_i$ (∵ u_i is base system of wieght space) — ①

$$\Rightarrow \quad C = \frac{1}{2}(\Sigma \alpha_i u_i^T) \cdot (\Sigma \alpha_i \lambda_i u_i)) \qquad (\because H u_i = \lambda u_i)$$

$$\Rightarrow \quad C = \frac{1}{2}(\Sigma \alpha_i^2 \lambda_i) \quad \Rightarrow \quad C = (\Sigma \alpha_i^2 \lambda_i) \quad \text{— ②}$$

From ①. $(w-w^*)(w-w^*)^T = \Sigma \alpha_i^2 \quad \Rightarrow \quad w_r \cdot w_r^* = \Sigma \alpha_i^2$

∴ from ① and ② we can state that they are ellipse with length propoti[onal]
to $1/\sqrt{\text{eigenvalues}}$

6) The issue with kaziranga National park is they have less amount of data to train a CNN so they should deploy the tegnique called <u>transfer-learing</u>

They should take the trained model of olympic national park and replace fc layers accordingly and train/tune the parameters of fc layer -s only.